JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 9, No. 1, June 2026, pp. 445~452
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v9i1.5638

Regression Model For Predicting Student Final Grades In Architecture And Computer Organization Courses

Gunardi Hamza[1], Nofri Wandi Al-Hafiz[2], Helpi Nopriandi[3], Febri Haswan[4]
[1,2,3,4]Teknik Informatika, Universitas Islam Kuantan Singingi

Article Info

Article history:

Received 05 01, 2026
Revised 06 11, 2026
Accepted 06 25, 2026

Keywords:

Multiple Linear Regression
Final Grade Prediction
Learning Analytics
Regression Model

ABSTRACT

The development of the digital era and the increasing need for human resources who are adaptable to information technology have prompted universities to utilize academic data to improve the quality of learning. This study aims to develop a multiple linear regression model to predict students' final grades in the Computer Architecture and Organization course based on learning evaluation variables. The predictor variables used include Quiz 1 score, Quiz 2 score, assignment score, Midterm Exam (ME) score, and Final Exam (FE) score, while the students' final grades are set as the dependent variable. The study employs a quantitative approach involving data collection, data preprocessing, splitting the dataset into training and testing sets, constructing the linear regression model, and evaluating model performance using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the coefficient of determination (R^2). The dataset was split into 75% training data and 25% testing data. The results indicate that adding predictor variables significantly improves model performance. The best model was obtained by combining quiz, assignment, and midterm exam variables, with an RMSE of 2.30, an MSE of 5.28, and an R^2 of 0.81. These findings indicate that multiple linear regression is capable of predicting students' final grades with a high degree of accuracy and can explain the relative contribution of each academic variable to students' learning outcomes. This study is expected to support the implementation of learning analytics and data-driven decision-making in the evaluation of learning at the university level.

Corresponding Author:

Gunardi Hamza
Program Studi Teknik Informatika
Universitas Islam Kuantan Singingi
Riau, Indonesia
Email: the.gun41@gmail.com

1. Introduction

The rapid development of the digital era, coupled with the increasing need for human resources capable of adapting to information technology, places higher education in a challenging position in its efforts to improve the quality of learning [1]. The utilization of academic data is one effective approach to identifying student learning patterns and supporting the development of evidence-based learning processes [2]. Through educational data analysis, institutions can conduct continuous evaluation and design improvements to student learning outcomes [3]. Thus, the management and utilization of academic data play a crucial role in supporting data-driven educational decision-making. Analytical approaches such as Educational Data Mining (EDM) and Learning Analytics are increasingly being applied in higher education as academic data becomes more abundant [3].

These approaches specifically focus on the early identification of potential learning issues through the prediction of student academic performance [4]. This prediction of academic performance plays a crucial role in designing more personalized and targeted learning interventions [5]. Thus, the prediction of student academic performance is an essential element in the development of effective adaptive learning systems.

Regression is one of the methods that has been widely applied to model and predict student learning outcomes, encompassing both statistical and machine learning approaches [6].

To illustrate the relationship between academic variables such as quiz scores, midterm exam scores, and other factors and students' final grades, linear regression is recognized as a technique that is simple, easy to understand, and effective [7]. The primary advantage of linear regression lies in its ability to explicitly explain the relative contribution of each predictor variable [8]. This characteristic makes linear regression an appropriate technique for predictive analysis in an educational context.

Given the conceptual and applied nature of the relevant courses, monitoring and evaluating student learning outcomes has become increasingly crucial in the fields of computer science and information technology [9].

The course on computer architecture and organization serves as a fundamental foundation for students to master the basic concepts of computing, information technology, and digital literacy. Students' academic performance in this course is often influenced by their success in advanced computing courses [10]. Consequently, the evaluation and prediction of learning outcomes in this foundational course are of great significance.

In the Computer Architecture and Organization course, assessment instruments such as quizzes, Midterm Exams (ME), and Final Exams (FE) are conventionally used to measure student learning outcomes [11]. Each of these evaluation components contributes differently to the calculation of students' final grades, while also reflecting various dimensions of conceptual understanding [12].

However, not all educational institutions utilize this assessment data for predictive analysis [13]. Therefore, an analytical model is needed that can systematically integrate evaluation elements to predict students' final grades. In the field of education, the regression approach has demonstrated high effectiveness in predicting student academic achievement through the analysis of learning assessment variables[14]

Furthermore, the predictive output from this model enables educators to identify quantitative components that have a dominant influence on student learning outcomes [15]. Thus, regression serves as a strategic analytical tool that significantly contributes to enhancing the quality of the learning process.

This study utilizes variables such as quiz scores, Mid-Semester Exam (MSE) scores, assignment grades, and Final Exam (FE) scores to develop a regression model for predicting students' final grades in the Computer Architecture and Organization course. The resulting model is expected to elucidate the causal relationships among evaluation variables and the accuracy of final grade predictions [16].

The lower the RMSE value produced by a predictive model, the higher the model's accuracy in generating estimates that are close to the actual values. This is because the magnitude of prediction error is inversely proportional to the Root Mean Squared Error (RMSE) value the smaller the error, the lower the measured RMSE value[17]. A key characteristic of RMSE is its high sensitivity to the presence of outliers or large-scale errors in the dataset. Due to this sensitivity, the RMSE metric is highly suitable for use as an evaluation indicator to measure the accuracy of predictive models in estimating students' academic performance[18].

Root Mean Squared Error (RMSE) is an evaluation metric used to measure the magnitude of the deviation between observed values and the predicted values generated by a regression model. The RMSE value is calculated by squaring the difference between each actual and predicted value, averaging the resulting squares, and then taking the square root [19].

Mean Squared Error (MSE) is one of the evaluation metrics used to assess the performance of a predictive model. Computationally, this metric is determined by calculating the average of the squared differences between the observed empirical values and the estimated values produced by the model. A low MSE value indicates that the model's ability to generate predictions is highly accurate relative to the actual values[20].

The coefficient of determination (R^2) is a statistical indicator used to measure the ability of a regression model to explain the proportion of variability in the dependent variable that can be accounted for by the independent variables collectively[21]. Numerically, the R^2 value falls within a closed interval ranging from 0 to 1. A regression model is considered to have optimal fit to the data if its R^2 value tends toward 1, indicating the model's superior ability to map or describe the functional relationship among variables[22].

Multiple Linear Regression is a multivariate linear regression analysis method that enables researchers to establish a linear functional relationship between one dependent variable and two or more independent variables. This technique is widely applied in various research fields for forecasting purposes, analyzing relationships among variables, and supporting data-driven decision-making[23]

The performance of a predictive model is generally evaluated using several statistical metrics, namely Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the coefficient of determination (R^2). RMSE is used to measure the average magnitude of prediction errors in the same units as the actual data, so a lower RMSE value indicates a higher level of model accuracy. Meanwhile, MSE evaluates the average square of the difference between the predicted and actual values, where larger errors incur a higher penalty. Additionally, the coefficient of determination (R^2) is used to measure the ability of independent variables to explain the variability of the dependent variable. A higher R^2 value indicates that the model is better at explaining the variation in students' final grades[24].

This study is anticipated to support faculty members and educational institutions in utilizing academic learning analytics data for predictive analysis [25], thereby contributing to the optimization of *learning analytics* implementation in learning evaluation at the university level.

2. Research Method

This study employs a quantitative paradigm with a focus on the development and validation of a regression model to predict students' final grades in the Computer Architecture and Organization course. The research process was conducted in a phased and structured manner, encompassing data collection, data preprocessing, predictive model construction, validity testing, and model performance assessment. The research workflow design aims to produce a predictive model with minimal error, measured using *the Root Mean Squared Error* (RMSE) metric.

If the evaluation indicates that the model's performance is not yet optimal, the model is reconstructed using different combinations of predictors until a model with superior performance is achieved

Figure 1. Research Workflow

A. Data Collection

This study utilizes primary data sourced from student learning assessments in the Computer Architecture and Organization course over a full semester. The collected data includes students' academic records representing learning achievements throughout the course, including quiz scores, assignment grades, Mid-Semester Exams (MSE), and Final Semester Exams (FSE).

All of this data was collected directly from the course assessment information system and served as the primary foundation for the development of the prediction model.

B. Training Data and Testing Data

To develop an optimal prediction model, this study selected a number of academic variables considered to have a significant influence on students' final grades. The independent variables (*predictors*) used include quiz scores, assignment scores, Mid-Semester Exams (MSE), and Final Semester Exams (FSE), with students' final grades set as the dependent variable.

The data used comes from the academic records of students who have taken and completed the relevant courses. Prior to the modeling stage, the data underwent a comprehensive *preprocessing* process that included verification of completeness, score consistency, and readiness for analysis. Subsequently, the processed *dataset* was randomly split into *a* 75% *training set* for constructing the regression model and *a* 25% *testing set* for evaluating the model's performance.

C. Linier Regression Model

This study applies the *linear regression* method to model the linear relationship between the dependent and independent variables. *Linear regression* focuses on determining the optimal coefficients that minimize *the* squared *error* between actual observed values and predicted values. The estimation of regression coefficients is performed using the *Ordinary Least Squares* (OLS) method, which is the standard approach in regression analysis.

The linear regression model applied can take the form of simple linear regression or multiple linear regression, depending on the number of predictor variables included. Mathematically, multiple linear regression is expressed through an equation that relates the dependent variable to multiple independent variables. In this study, the dependent variable (Y) is defined as students' final grades in the Computer Architecture and Organization course, while the independent variables include:

X1 = Quiz 1 Score
X2 = Quiz 2 Score
X3 = Assignment Score
X4 = Midterm Exam Score
X5 = Final Exam Score

D. Model Evaluation

The performance of the regression model is evaluated using the *Root Mean Squared Error* (RMSE) metric, which measures the average magnitude of the model's prediction error relative to the actual observed data. The RMSE value is calculated as the square root of the mean *squared error*, which is the difference between the predicted value and the observed value, using the following mathematical formula:

a. Root Mean Squared Error (RMSE)

$$\mathbf{RMSE = \sqrt{\left(\Sigma(yi - \hat{y}i)^2 / n \right)}}$$

Notes :
yi = actual value/observed value
ŷi = predicted values
n = amount of data
I = sequence data

b. Mean Squared Error (MSE)

$$\mathbf{MSE = \Sigma(yi - \hat{y}i)^2 / n}$$

Notes :
yi = actual value
ŷi = predicted values
n = amount of data

c. Coefficient of Determination (R^2)

$$\mathbf{R^2 = 1 - (SS_{res} / SS_{tot})}$$

Notes :
SS_{res} = sum of squared residuals (errors)
SS_{tot} = sum of squares

A lower RMSE value reflects the model's superior predictive capability regarding students' final grades. This study developed several variants of regression models with different combinations of predictors, which were then evaluated comparatively using the RMSE metric. The model with the lowest RMSE value was selected as the optimal model because it demonstrated the highest level of prediction accuracy.

3. Result and Discussion

This study successfully developed four variants of multiple linear regression models with different combinations of predictors, aiming to identify the optimal model for predicting students' final grades in the Computer Architecture and Organization course.

The performance of each model was assessed using the *Root Mean Squared Error* (RMSE), *Mean Squared Error* (MSE), and coefficient of determination (R^2) metrics. A comparative analysis of these models revealed the impact of adding or reducing predictor variables on prediction accuracy.

A. Results of the Multiple Linear Model

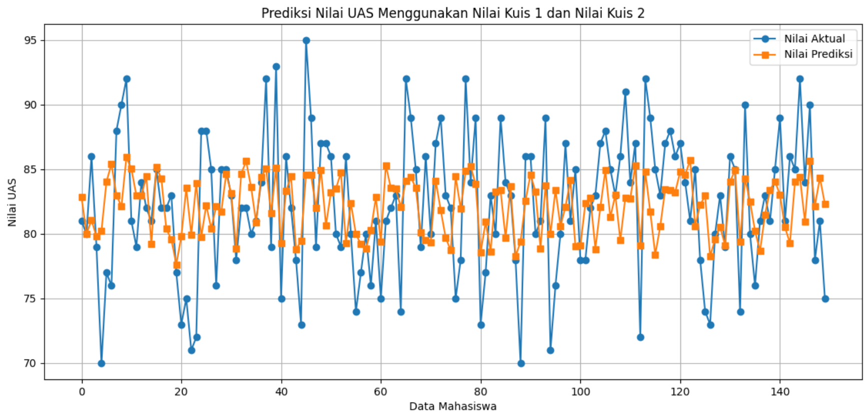

Figure 2. Grade Prediction with Model 1

Figure 2 illustrates Model 1, constructed based on two predictor variables: X_1 (Quiz 1 score) and X_2 (Quiz 2 score). Evaluation analysis revealed that the model achieved an RMSE of 4.42, an MSE of 19.57, and an R^2 of 0.29. The relatively high RMSE value coupled with a low R^2 indicates the model's limitations in capturing the variability of students' final grades. This finding confirms that these two predictors are insufficient to comprehensively represent all factors determining students' final grades.

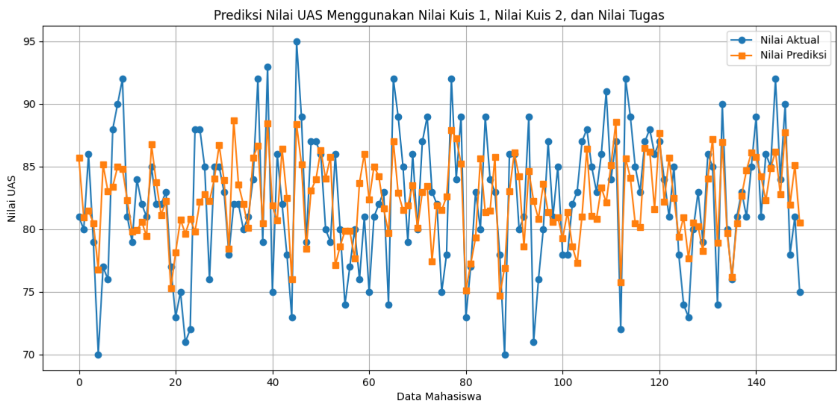

Figure 3. Grade Prediction with Model 2

Figure 3 presents Model 2, which was developed by adding one new predictor variable, X_3, thereby incorporating X_1, X_2, and X_3 as a whole. The evaluation results indicate an improvement in performance relative to Model 1, with an RMSE of 3.88, an MSE of 15.09, and an R^2 of 0.45. The reduction in the RMSE value and the increase in R^2 underscore the positive contribution of the additional predictor variable to the accuracy of final grade predictions, although this improvement remains marginal.

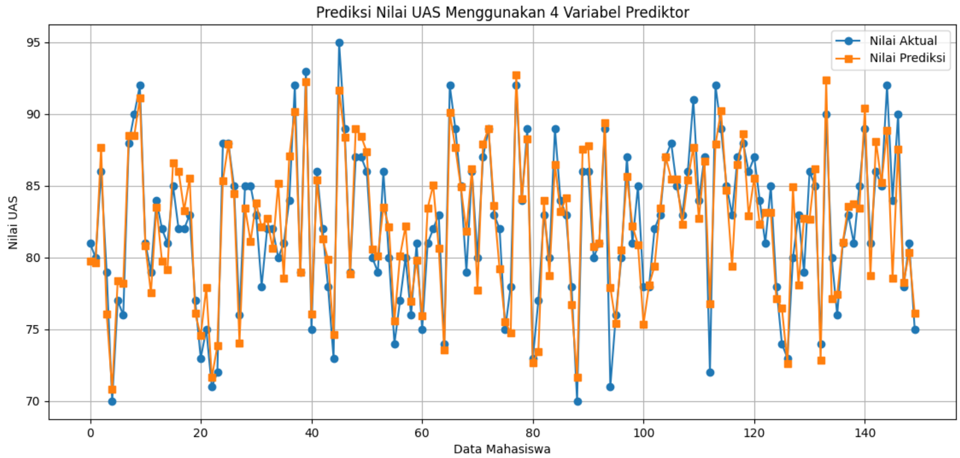

Figure 4. Grade Prediction with Model 3

Comprehensive set of predictors, including X_1, X_2, X_3, and X_4. This model yields an RMSE of 2.30, an MSE of 5.28, and an R^2 of 0.81. These results reflect a substantial improvement in prediction precision compared to Model 1 and Model 2. The substantial reduction in the RMSE underscores the crucial role of variable X_4 in explaining the variability in students' final grades. Overall, this model's performance demonstrates superior accuracy in predicting students' final grades.

B. Analysis of Model Prediction Results

After constructing a multiple linear regression model using Quiz 1 Score, Quiz 2 Score, Assignment Score, and Midterm Exam Score as predictor variables, the test data was evaluated to assess the model's ability to predict students' Final Exam Scores. The estimated values were then compared with the actual values and visually represented in the form of a scatter plot, as shown in Figure 5.

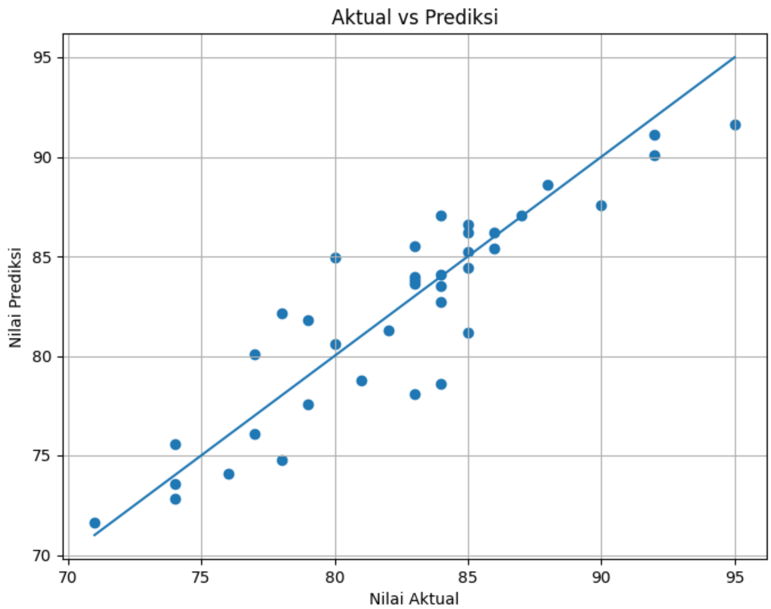

Figure 5. Actual vs Predicted

As shown in Figure 5, it can be observed that the majority of data points are concentrated around the diagonal line, which represents the ideal condition that is, a situation where the estimated values match the actual values. The density of the data points along this diagonal line indicates that the developed model is sufficiently capable of predicting students' final exam scores. In addition, the distribution pattern of the data points, which consistently follows the diagonal line, indicates a strong linear correlation between the actual values and the estimated values. Although there are a few points that deviate slightly above or below the reference line, these deviations remain within acceptable limits and therefore do not indicate statistically significant prediction errors.

The results of this visual representation indicate that the developed regression model is capable of adequately capturing the relationship between Quiz 1 Score, Quiz 2 Score, Assignment Score, and Midterm Exam Score and the Final Exam Score. The proximity of the data points to the diagonal line is directly proportional to the accuracy of the estimates produced by the model; the smaller the distance of the points from the line, the higher the precision of the predictions obtained.

Overall, the comparison graph between actual and estimated values shows that the multiple linear regression model is capable of producing predictions that closely approximate the actual observed values. This implies that the constructed model is sufficiently valid to be implemented as a supporting tool in predicting students' final exam scores.

4. Conclusion

Based on the findings of this study, it can be concluded that the application of learning assessment data enables multiple linear regression to effectively predict students' final grades in the Computer Architecture and Organization course. The inclusion of academic variables such as quiz scores, assignments, midterm exams, and final exams provides a quantitative representation of the relative influence of each element on students' final performance. An evaluation of the three regression models reveals that the composition of predictor variables substantially shapes model performance. Models relying on predictors with a more appropriate number and relevance tend to achieve higher coefficients of determination and lower prediction error rates. In particular, models incorporating summative evaluation components such as assignments and exams demonstrate superior performance in capturing predictive variation.

The results of the evaluation of the three regression models indicate that the composition of predictor variables significantly shapes model performance. Models utilizing predictors with more appropriate quantities and categories have the potential to achieve higher coefficients of determination and reduce prediction error rates. Specifically, models that integrate summative assessment elements such as lab work and exams demonstrate superiority in capturing prediction variations. The findings of this study confirm that regression analysis is not only useful for predictive purposes but also for evaluating the learning process. Therefore, regression models can be utilized as a foundation for academic decision-making, optimizing teaching strategies, and developing more effective evaluation systems.

Acknowledgement

The author would like to express appreciation to all students who participated in the data collection process for the Computer Architecture and Organization course. Thanks are also extended to colleagues, reviewers, and everyone who provided advice, feedback, and support during the preparation of this study. The author hopes that this study will contribute to the development of learning analytics and academic prediction models in higher education.

References

[1] K. Dadang, "TRANSFORMASI DIGITAL DALAM PENDIDIKAN: TANTANGAN DAN PELUANG MENUJU PEMBELAJARAN DI ERA DIGITAL," vol. 10, 2025.

[2] D. Silveti dan K. Andersson, "Challenges of Governing O ff -Grid ' Productive ' Sanitation in Peri-Urban Areas : Comparison of Case Studies in Bolivia and South Africa," 2019.

[3] U. Garut, N. A. Dwijaksana, dan A. H. Muhammad, "Pemanfaatan Teknologi Learning Analytics Dalam Mendeteksi Pola Belajar Untuk Meningkatkan Kualitas Pembelajaran di Smk Negeri 1 Kabupaten Sorong," hal. 122–128, 2025.

[4] R. Hasudungan dan W. J. Pranoto, "Implementasi Teorema Naïve Bayes Pada Prediksi Prestasi Mahasiswa," vol. 5, no. 1, hal. 10–16, 2021.

[5] A. Florent, M. Rizal, dan S. Oka, "Prediksi Kinerja Siswa SMP Berdasarkan Data Akademik dan Perilaku Menggunakan Machine Learning," vol. 18, no. 2, hal. 67–76, 2025.

[6] D. P. Wiguna, "Pemodelan Prediksi Mahasiswa Berprestasi Akademik Dengan Multi-Linear Regression (Studi

Kasus : Prodi Sistem Informasi) Complex : Jurnal Multidisiplin Ilmu Nasional," vol. 2, hal. 53–58, 2025.
[7] R. Rismaya, D. Yuniarto, dan D. Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," 2025.
[8] Khoirul Bariah Rambe, "PEMANFAATAN REGRESI LINIER SEDERHANA DALAM PENELITIAN PENDIDIKAN DASAR: KAJIAN LITERATUR SISTEMATIS," vol. 10, 2025.
[9] M. C. Andi, Jumardi, P. Sejarah, dan H. Artikel, "EVALUASI PEMBELAJARAN BERBASIS TEKNOLOGI INFORMASI DAN KOMUNIKASI," vol. 19, hal. 1–5, 2020.
[10] D. A. Septriansyah, R. R. Darmawan, dan R. Y. Pradana, "Penerapan AI dalam Mata Kuliah Arsitektur dan Organisasi Komputer : Studi Kasus Mahasiswa Pendidikan Ilmu Komputer Universitas Pendidikan Indonesia," vol. 8, hal. 49174–49181, 2024.
[11] B. A. Indriasari, U. K. Musi, M. R. Efila, dan U. K. Musi, "ANALISIS VALIDITAS SOAL UJIAN AKHIR SEMESTER MATA KULIAH WAJIB UMUM (MKWU) BAHASA INDONESIA," vol. 12, hal. 20–26, 2021.
[12] D. Dari dan R. Kognitif, "ANALISIS PEMAHAMAN KONSEP MAHASISWA PADA MATERI ANALISIS REAL BERDASARKAN TAKSONOMI BLOOM," vol. 7, no. 2, hal. 153–163, 2020.
[13] Y. N. Sukmaningtyas, R. M. Akbar, dan G. Rohma, "Penerapan Predictive Analytics untuk Analisis Faktor-faktor yang Mempengaruhi Performa Akademik Siswa menganalisis faktor-faktor yang mempengaruhi performa akademik siswa telah," vol. 4, no. 2, hal. 127–145, 2024.
[14] F. Riandari, H. T. Sihotang, dan H. Husain, "Forecasting the Number of Students in Multiple Linear Regressions," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, hal. 249–256, 2022, doi: 10.30812/matrik.v21i2.1348.
[15] W. C. Ferdinan, M. R. Noerfikri, dan P. A. Panchadri, "Implementasi Algoritma Regresi Linear Berganda untuk Memprediksi Prestasi Siswa," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2228.
[16] B. Rizky, S. Dwinanda, A. N. Ilhami, H. Choerunisya, dan I. J. Thira, "Jurnal Sistem dan Teknologi Informasi Indonesia Prediksi Nilai Akhir Mahasiswa Dengan Metode Naïve Bayes Prediction of Final Student Grades Using Naive Bayes Method," vol. 9, no. 1, hal. 27–35, 2024.
[17] P. Ann, R. Forest, D. A. N. Regresi, M. R. Nugraha, H. A. Rusdan, dan M. Sinambela, "PREDIKSI EVAPORASI BERBASIS MESIN :," vol. 9, no. 2, hal. 380–386, 2025.
[18] P. Curah Hujan Dengan Validasi Perbandingan Metode Backpropagation Dan *et al.*, "Partial Least Square Regression," vol. 6, hal. 292–301, 2021.
[19] K. Farouq Mauladi, "Perbandingan Metode Regresi Linear Dan Neural Network Backpropagation Dalam Prediksi Nilai Ujian Nasional Siswa Smp Menggunakan Software R," *Joutica*, vol. 5, no. 1, hal. 331–336, 2020.
[20] D. N. Huda dan S. Handayani, "Prediksi Nilai Ujian dengan Artificial Neural Network," *Remik*, vol. 7, no. 1, hal. 157–165, 2023, doi: 10.33395/remik.v7i1.11983.
[21] S. Aminuddin, Sudarno, "PEMILIHAN MODEL REGRESI LINIER MULTIVARIAT TERBAIK DENGAN KRITERIA MEAN SQUARE ERROR," vol. 2, hal. 11–18, 2013.
[22] D. Darnah, "Menentukan Model Terbaik dalam Regresi Poisson dengan Menggunakan Koefisien Determinasi," *JMSK J. Mat. Stat. dan Komputasi*, vol. 6, no. 2, hal. 59–71, 2010.
[23] A. Rahmadhani, D. S. Sihabudin Sahid, dan Y. D. Lulu Widyasari, "Implementasi SEM-Multiple Linear Regression dalam Prediksi Jumlah Pendaftaran Mahasiswa Baru di Perguruan Tinggi XYZ," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, hal. 150–162, 2023, doi: 10.25077/teknosi.v9i2.2023.150-162.
[24] A. Qoiriah dan Y. Yamasari, "Prediksi Nilai Akhir Mahasiswa Dengan Metode Regresi (Studi Kasus Mata Kuliah Pemrograman Dasar)," *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 1, hal. 40–43, 2021, doi: 10.26740/jieet.v5n1.p40-43.
[25] P. S. Saputra *et al.*, "ANALISIS PREDIKTIF DROPOUT MAHASISWA BERDASARKAN KINERJA AKADEMIK SEMESTER AWAL MENGGUNAKAN," vol. 07, no. 01, hal. 164–171, 2026.